SNPStudio™

Version 1.0

User Manual

ISTECH
Beyond Bioinformatics

First of all,

Thank you for purchasing our product.

This User Manual contains information about SNPStudio version 1.0.
It describes notable information in using, installation instructions and operating methods.
Please read the manual before using to manage the program properly.

• This User Manual and the product may have technical error, and this can be adjusted without pre-announcement.
• We will not take any responsibility of the damage caused from using this product beside the purpose and the method described in this User Manual.

To use this product properly and safely, please read this User Manual thoroughly before using. Thank you.

<Customer Support>
Istech Corp. is always listening to our customers. If you have any questions, please feel free to contact us as shown below. Thank you.
Tel No : 82-31-903-1155
Fax No : 82-31-903-1152
e-mail : snpstudio@istech21.com
Website : http://snp.istech21.com/snpstudio
Technical Support: yblee@istech21.com

Table of Contents

Figure Legends

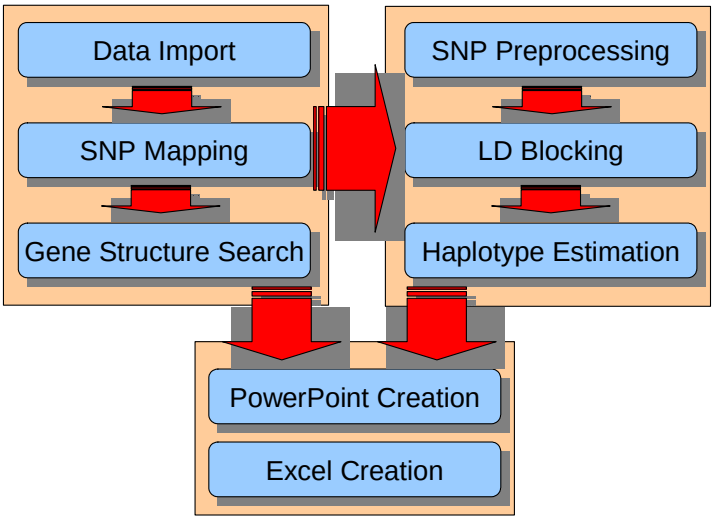
Chapter 1

Introduction

1. Introduction

1.1. Summary

SNPStudio is a software program which transforms SNP information into a PowerPoint OLE object with associated genome and gene structures. And, if there are genotype information, SNPStudio automatically performs various analyses like calculation of linkage disequilibrium (LD) among adjacent SNPs, construction of LD blocks and haplotype estimation within the constructed LD block. It is possible to analyze up to tens of thousands SNPs and whole analysis procedure is controlled in detail through SNPStudio run wizard. The analyzed results are automatically created into OLE objects and easy to edit/manipulate for visual enhancements within PowerPoint.



<Figure 1-1> Operating Process of SNPStudio

1.2. Main Features

1.2.1. Input Data Type

SNPStudio supports various input formats as followings:

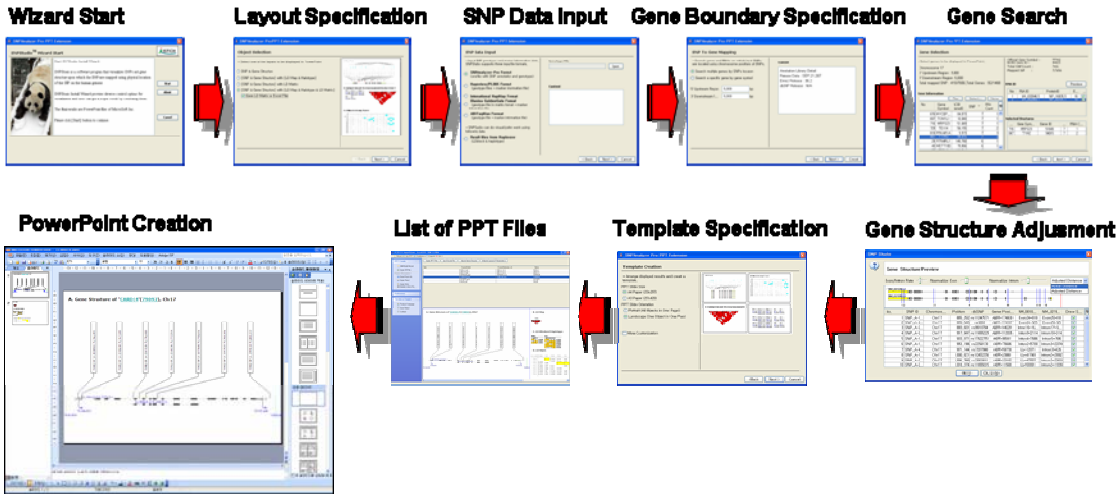
- Haploview / PLINK Format (genotype file and marker information file)
- International HapMap Genotype Format
- Illumina GoldenGate Format (genotype file in matrix format and marker information file)
- ABI TaqMan Format (multiple genotype files and one marker information file)
- Result files from Haploview
- SNPAnalyzer-Pro Format (one file containing SNP annotation and genotype)

※ Maximum number of SNPs for analysis: Over 50,000

※ Maximum number of samples for analysis: Over 2000

1.2.2. Run Wizard

SNPStudio operates all analyses through run wizard from input data to gene searching, SNP mapping, and visualization and analysis parameter setup.



<Figure 1-2> Operation Process of Run Wizard

1.2.3. Gene Searching and SNP Mapping

SNPStudio searches all the genes on which SNPs in the input data are located using chromosome position of SNPs and Ref. Seq. information provided by NCBI. Followings are the information used for gene searching and SNP mapping.

- Official Gene Symbol
- NCBI Gene ID
- Start/Stop Position of a Gene
- RNA Count of a Gene
- RNA ID(s)
- Protein ID(s)
- Chromosome position of SNPs

1.2.4. Linkage Disequilibrium Analysis

If the input data contains individuals' genotype information, SNPStudio automatically performs linkage disequilibrium analyses. The contents of the linkage disequilibrium analysis are as follows.

- Measuring Linkage Disequilibrium between Adjacent SNPs
- Constructing Linkage Disequilibrium Blocks (LD Blocks)
- Estimating Haplotypes in Each LD Blocks
- Pairwise Tagging SNPs Selection
- Haplotype Tagging SNPs Selection

1.2.5. Creation of OLE Object of PowerPoint

All the analyzed results are created as OLE objects of PowerPoint of MicroSoft and easy to edit/manipulate for visual enhancements within PowerPoint. Followings are the types of OLE objects.

- SNP Information
  - SNP ID, dbSNP #rs, Chromosomal Position and etc.
- Gene Information
  - Gene Symbol, NCBI Gene ID, Gene Orientation
  - mRNA ID, Protein ID
  - Exon, Intron, UTR
  - Etc.

- Linkage Disequilibrium Information
  - Linkage Disequilibrium Map/Block/Matrix
  - Haplotypes in LD Blocks
  - Pairwise Tagging SNPs
  - Haplotype Tagging SNPs
  - Etc.

1.3. System Requirement

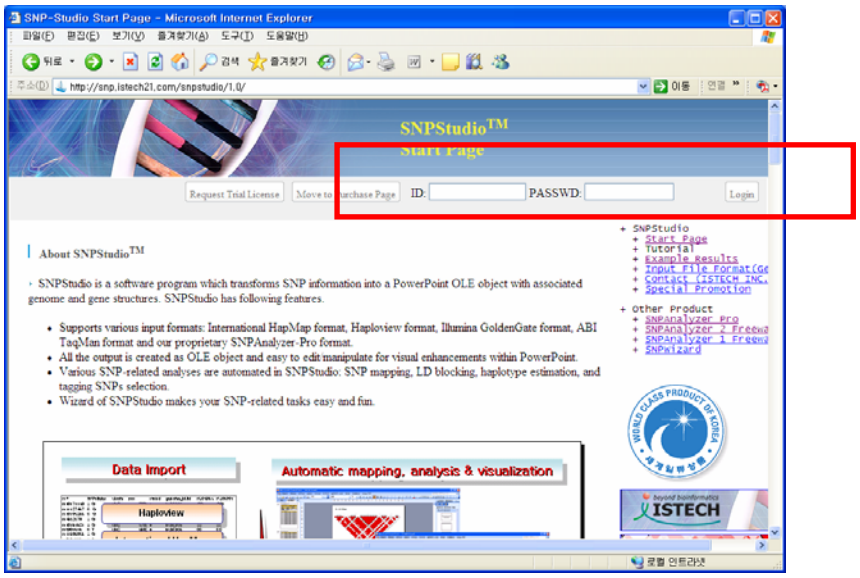
- Minimum Specifications
  - OS: Microsoft Windows 2000/XP/Vista System (internet connection required)
  - CPU: Pentium 4 2.4GHz or higher
  - RAM: 1GB or more
  - Web Browser: Internet Explorer 6.0 or higher
- Required Application Program
  - J2SE Runtime Environment 5.0 or higher (installed with SNP Analyzer-Pro)
  - MicroSoft PowerPoint 2003 or higher

Chapter 2

SNPStudio Operation

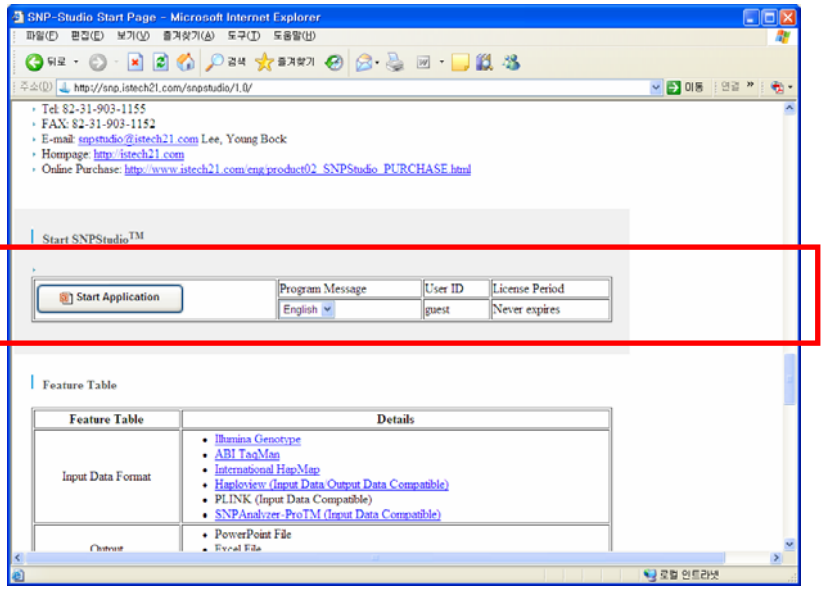
2. Login and Start SNPStudio

SNPStudio is a software program which operates on web browser after downloading execution file. <Figure 2-1> shows the homepage of SNPStudio, and the application program can be implemented after login with your ID and password.



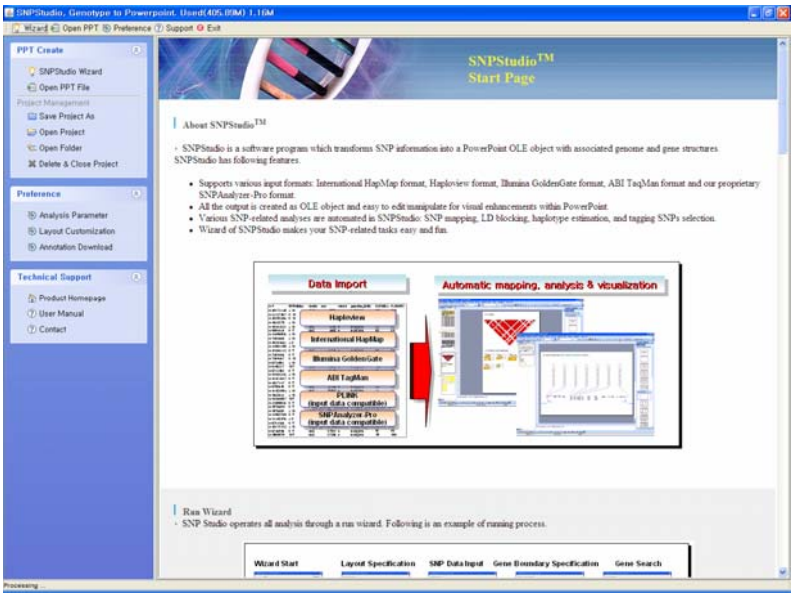
<Figure 2-1> SNPStudio Login Page

After login, scroll down the page and click [Start Application] button as shown in <Figure 2-2>, then it triggers application interface of SNPStudio as shown in <Figure 2-3>.



<Figure 2-2> SNPStudio Start Button

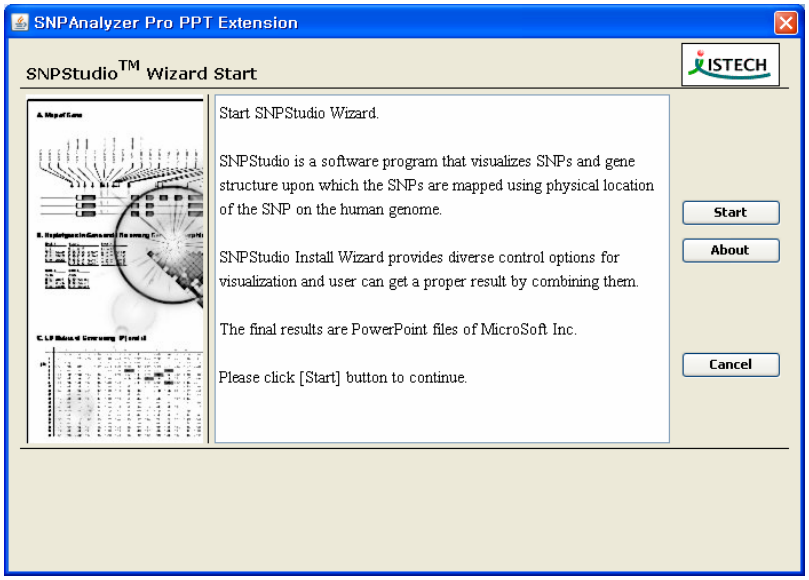
On the upper side in the application interface, there are hot keys like [Wizard], [Open PPT], [Preference], [Support] and [Exit]. On the left side of the interface, there are menus required to operate and manage the program.



<Figure 2-3> Application Interface of SNPStudio

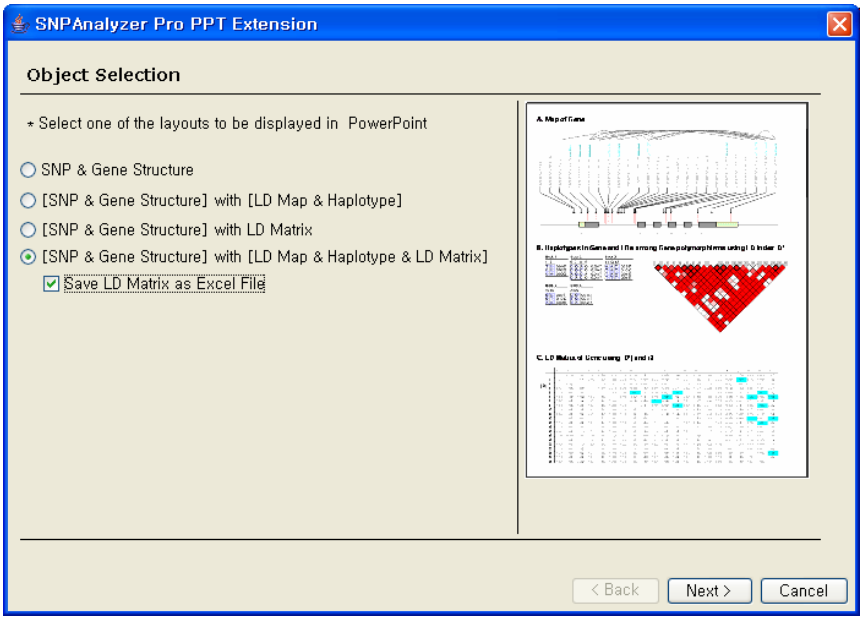
2.1. Run Wizard

SNPStudio always starts from run wizard. Click [SNPStudio Wizard] on the left side of the application interface, or click the hot key [Wizard] on the top of the interface, then a window of run wizard like <Figure 2-4> pops up. Click [Start] button to go to the next step.



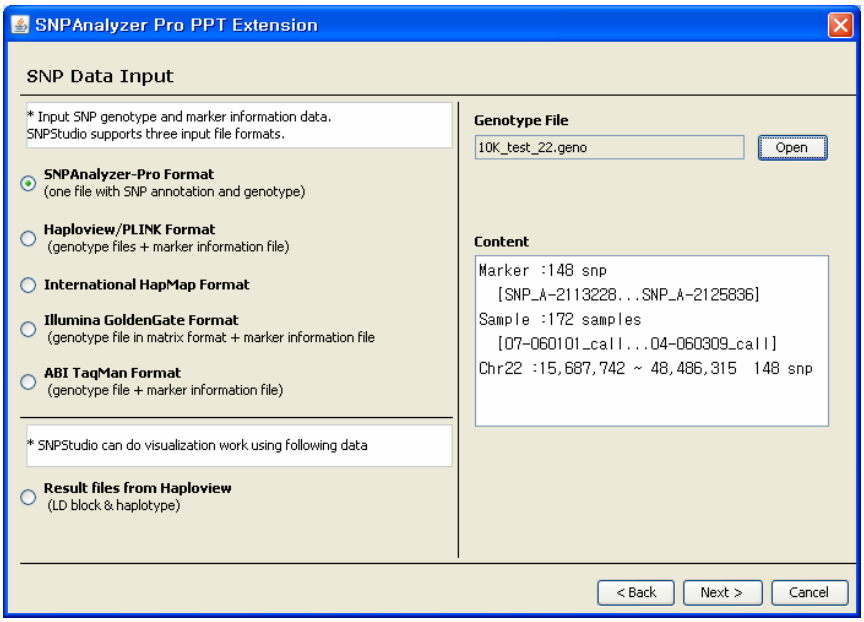
<Figure 2-4> Wizard Window of SNPStudio

On this step, it is possible to select layout configuration from four different options as shown in the <Figure 2-5>. Selecting each item, you can see the example result on the right side of the interface. Click [Next] button to go to the next step.



<Figure 2-5> Layout Configuration of OLE Objects

On this step, you can select one input data format among six different formats as shown in the <Figure 2-6>. Select relevant data format and click [Open] button to input your data. When the data input is done, click [Next] button to go to next step.



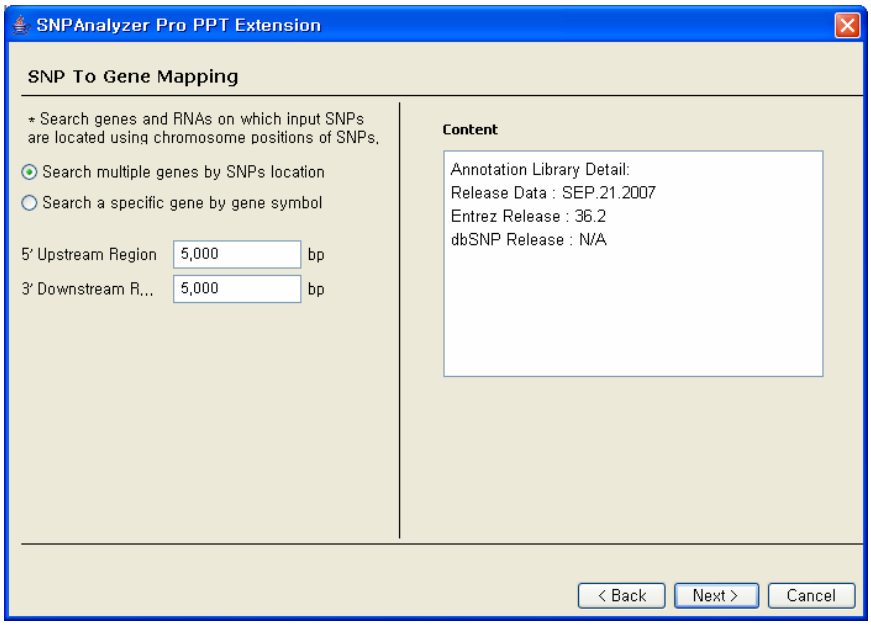
<Figure 2-6> SNP Data Input

On this step, it is possible to set up searching method of the genes in addition to SNP mapping. Followings show details of the searching method.

- Search multiple genes by SNPs location: search for all the genes on which SNPs in the input data are located.
- Search a specific gene by gene symbol: search for a specific gene using gene symbol or NCBI gene ID.

The default values of [5' Upstream Region] and [3' Downstream Region] are 5,000 bp and these mean that the upstream and downstream boundaries of the genes on which SNPs in the input data are located are 5,000 base pairs each.

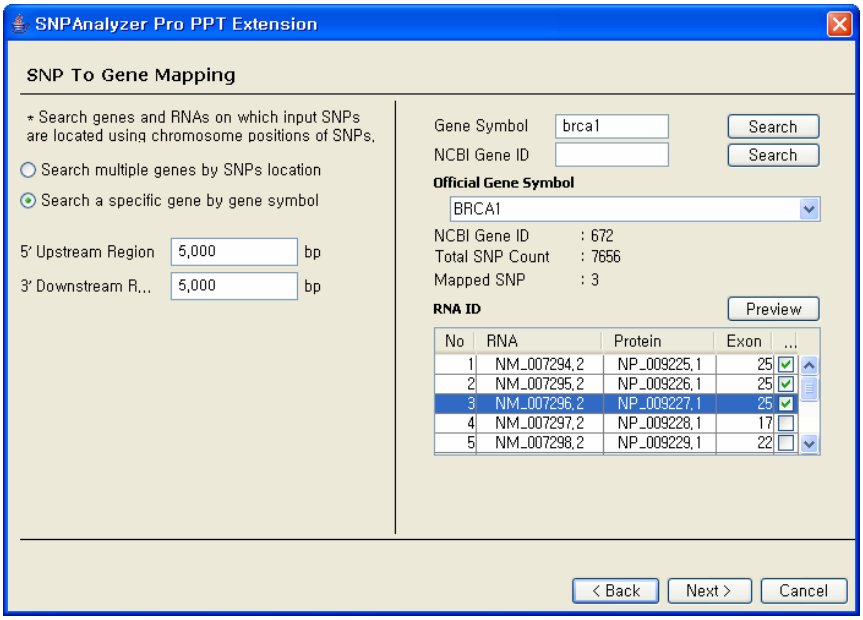
<Figure 2-7> is the interface to search for all the genes on which SNPs in the input data are located. Click [Next] button to go to next step



<Figure 2-7> Search Multiple Genes by SNPs Location

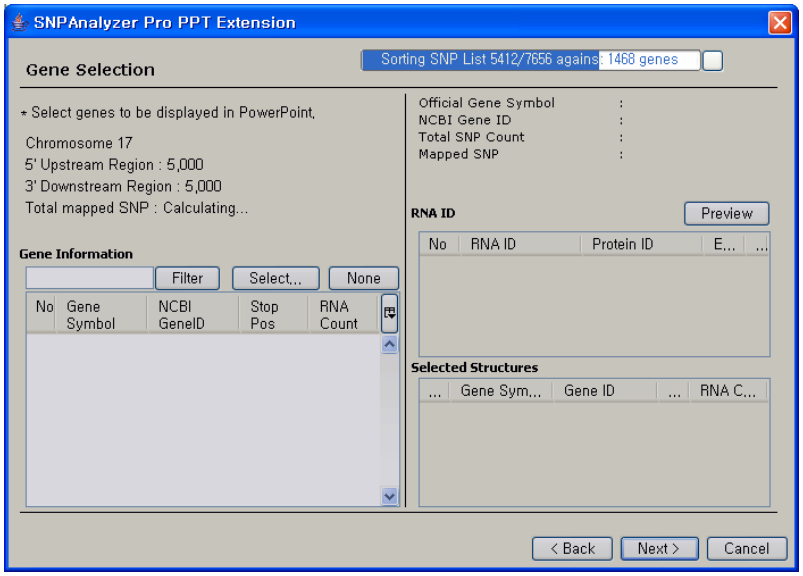
.

<Figure 2-8> is the interface to search a specific gene using gene symbol or NCBI gene ID. Input gene symbol or NCBI gene ID and click [Search] button to see the table with the associated information such as RNAs and proteins. Select RNAs using check box to be created as OLE objects. Click [Next] button to go to next step.



<Figure 2-8> Search a Specific Gene by Gene Symbol

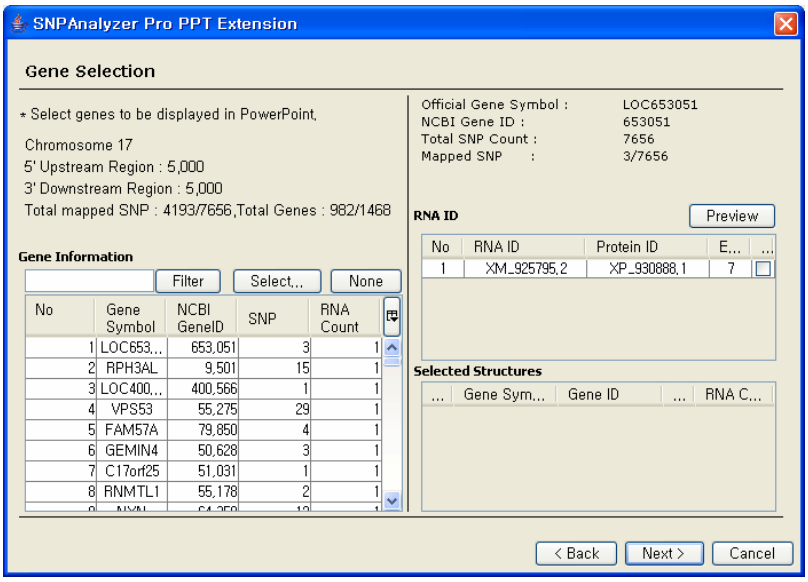
<Figure 2-9> shows that the searching and mapping process is running by the [Search multiple genes by SNPs location] method.



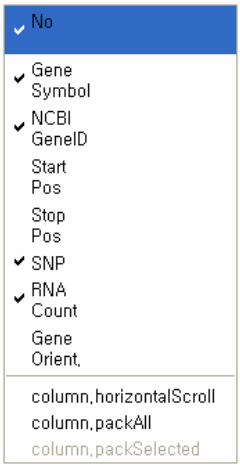
<Figure 2-9> Searching and Mapping Status

<Figure 2-10> is the result of searching and mapping. The contents in the table in the left side of the window are gene symbol, gene ID, number of RNAs and number of SNPs located in the relevant genes. You can sort the contents in the table by clicking header of columns. Click 

button on the right side of the table, then you can add or remove the contents in <Figure 11> to the table by checking or unchecking.



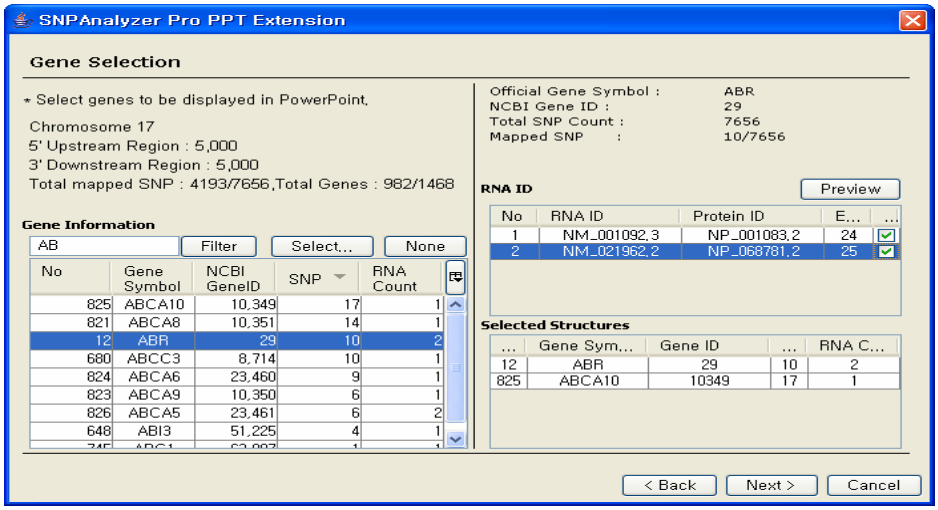
<Figure 2-10> Searching Result



<Figure 2-11> List of Gene Relevant Contents

As seen on the <Figure 2-12>, input “AB” on the empty text box right below the “Gene Information” title, then genes with gene symbols starting with “AB” only will be listed in the table. Select the gene that you want to create as OLE object, then the relevant RNA IDs and protein IDs will be listed on the right side of the window. Select RNAs in the list using check box, then these will be created as OLE objects of PowerPoint. If you select gene in the list titled as “Selected Structures” and uncheck RNAs in the list titled as “RNA ID”, then you cancel the creation of OLE objects about the relevant gene. You want all the genes listed in the left table to

be created as OLE objects, just click [Select All] button. Click [None] to cancel the checking status. [Preview] button enables user to preview and modify the SNPs locations and gene structures in a separate window as shown in <Figure 2-13>. Click [Next] button to go the next step.

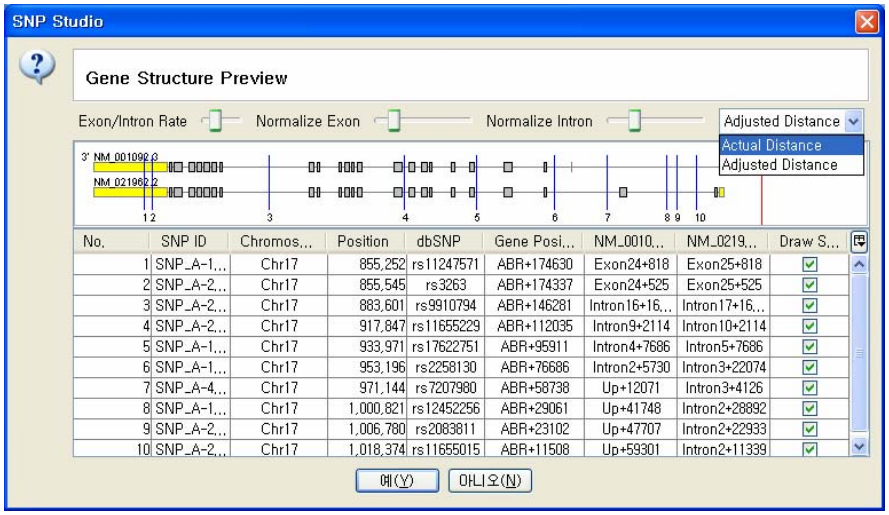


<Figure 2-12> Gene Filtering and Gene Selection

In the <Figure 2-13>, you can modify SNPs locations and gene structures using four configuration methods:

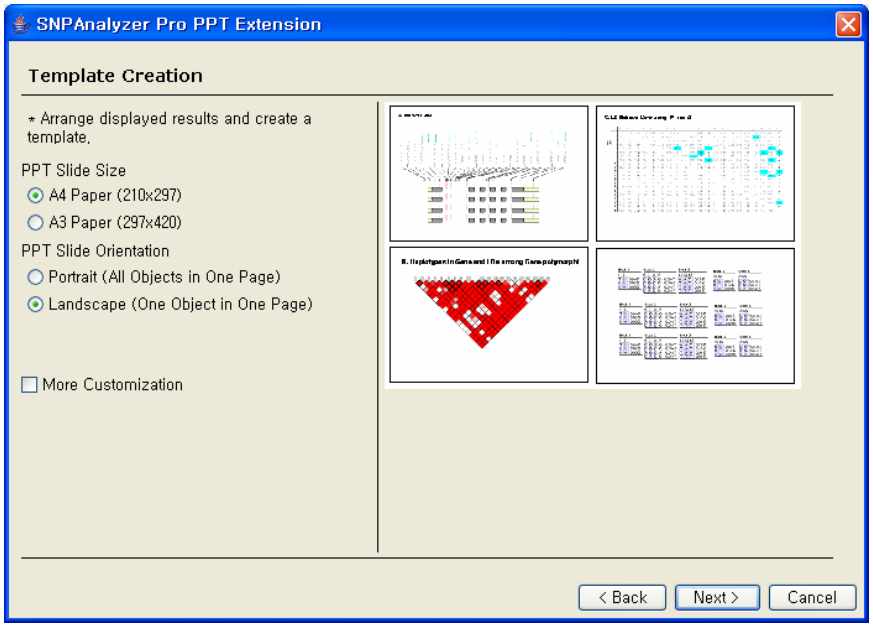
- Exon/Intron Rate: Control the ratio of the length of exon and intron
- Normalize Exon: Control the relative lengths of exons regarding other exons' lengths
- Normalize Intron: Control the relative lengths of exons regarding other introns' lengths
- Actual Distance/Adjusted Distance

If you do not want to control the ratio of the length of Exon/Intron, select [Actual Distance] option in the list of the right side of the window. Default value is [Adjusted Distance]. Preview of gene structures and locations of SNPs are displayed right below the control panel. You can add or remove SNPs for visualization using check box. Click [Yes] button to apply the modified configuration and proceed to the next step.

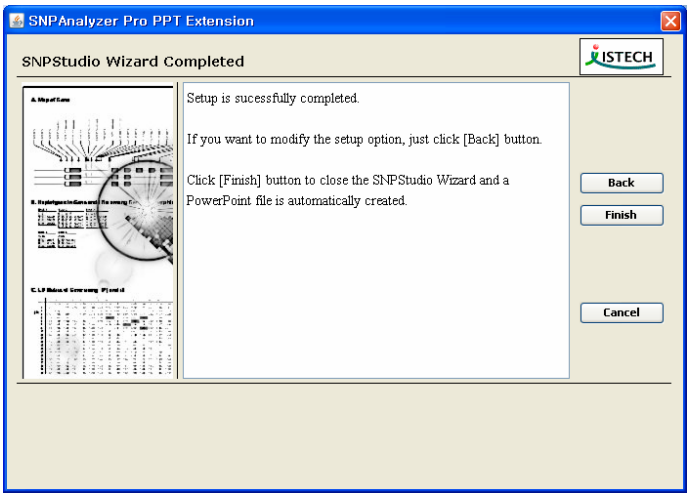


<Figure 2-13> Preview of Gene Structures and SNPs Locations

In this step, as shown on the <Figure 2-14>, you can configure page size and orientation of PowerPoint slide. Default configuration is “A4 size” and “Landscape”. When the set up is done, click [Next] button, then you can see the completion message of run wizard as shown in the <Figure 2-15>. Click [Finish] button and it starts the creation of OLE objects based on the previously setup configurations.



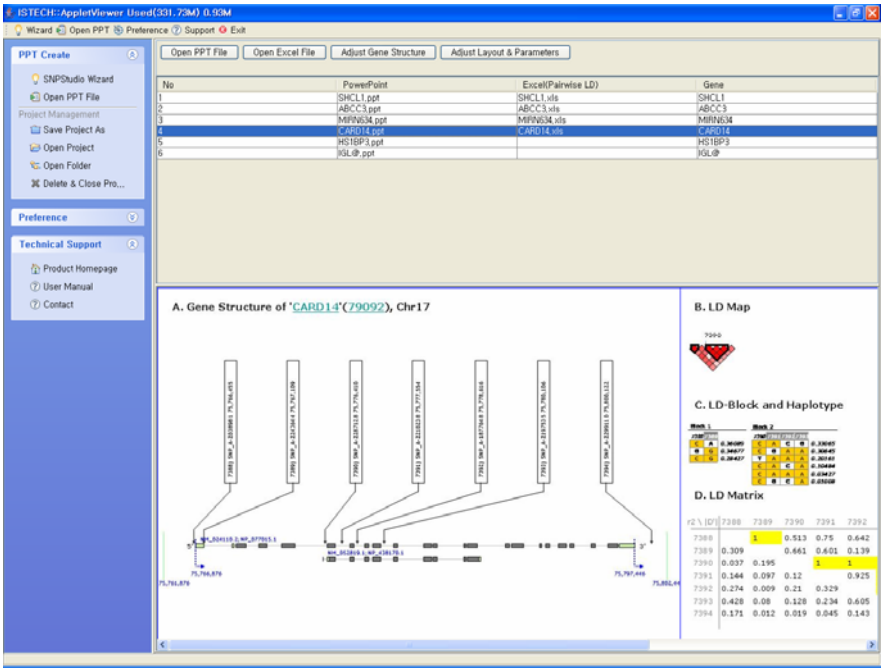
<Figure 2-14> PowerPoint Template Creation



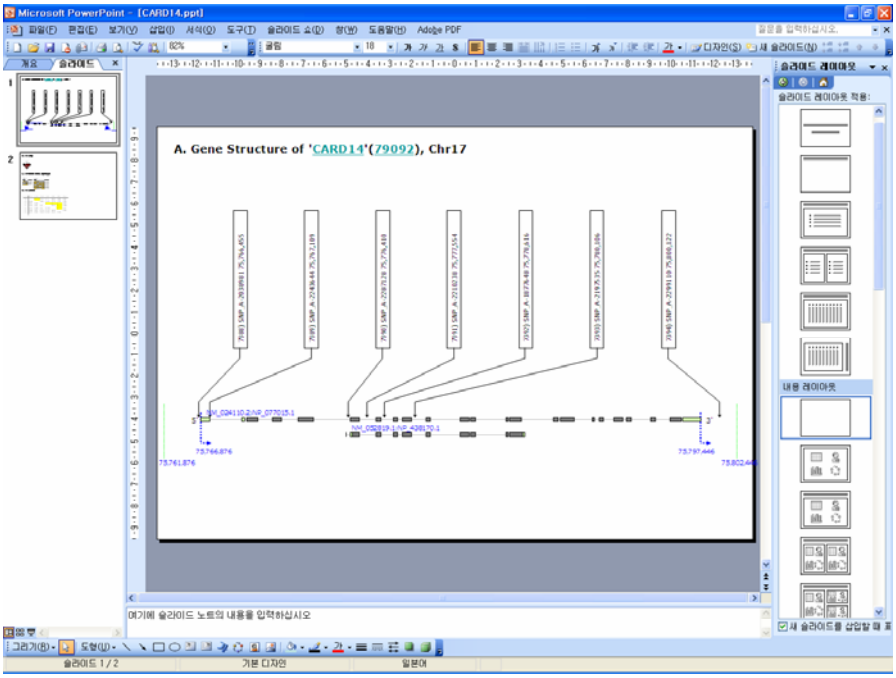
<Figure 2-15> Completion Message of Run Wizard

2.2. Operation Result Confirmation

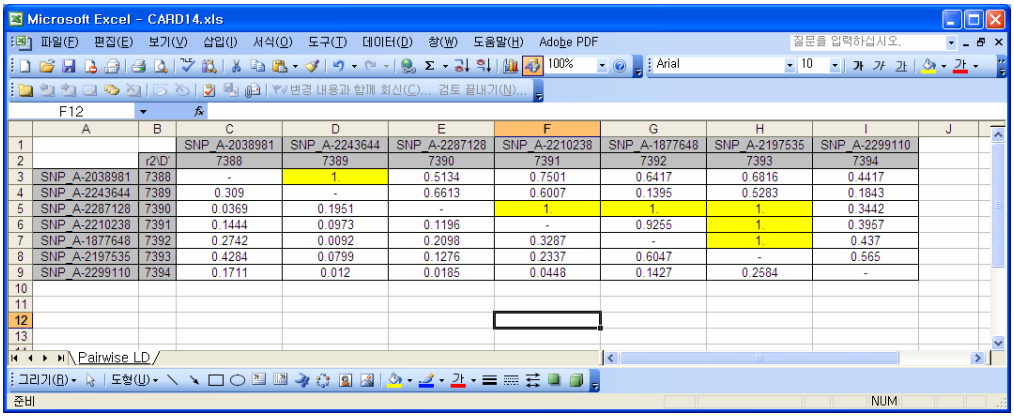
When the run wizard operation is finished, you will have the result as seen on the <Figure 2-16>. On the upper side of the window is the list of created OLE objects of PowerPoint. Select one from the list and click [Open PPT File] button to activate PowerPoint Program. Then you will see the visualization result as seen on the <Figure 2-17>. Click [Open Excel File] button to activate Excel program and you will have the linkage disequilibrium analysis result (D' and R^2) as you can see on the <Figure 2-18>.



<Figure 2-16> List of OLE objects and Summary of Display



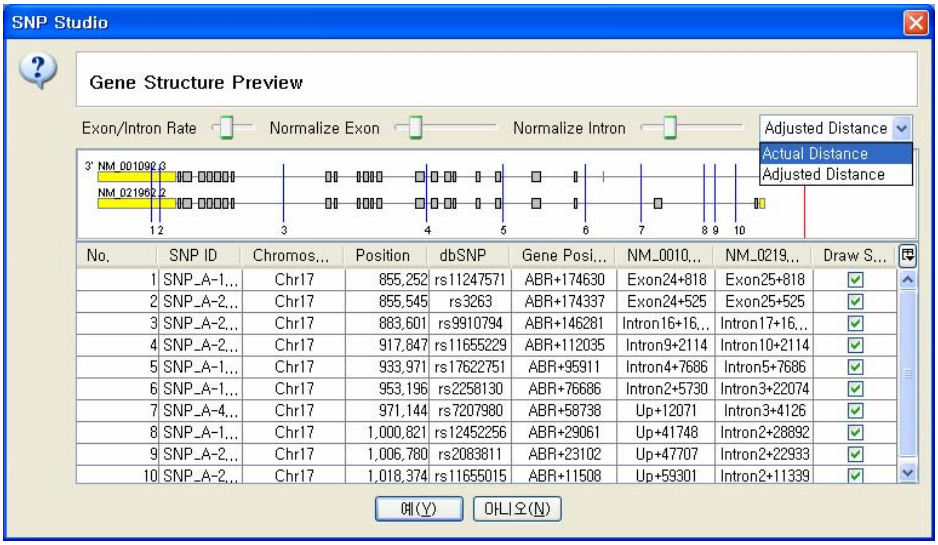
<Figure 2-17> PowerPoint Display



	A	B	C	D	E	F	G	H	I	J
1			SNP_A-2038981	SNP_A-2243644	SNP_A-2287128	SNP_A-2210238	SNP_A-1877648	SNP_A-2197535	SNP_A-2299110	
2		r2\D'	7388	7389	7390	7391	7392	7393	7394	
3	SNP_A-2038981	7388	-	1.	0.5134	0.7501	0.6417	0.6816	0.4417	
4	SNP_A-2243644	7389	0.309	-	0.6613	0.6007	0.1395	0.5283	0.1843	
5	SNP_A-2287128	7390	0.0369	0.1951	-	1.	1.	1.	0.3442	
6	SNP_A-2210238	7391	0.1444	0.0973	0.1196	-	0.9255	1.	0.3957	
7	SNP_A-1877648	7392	0.2742	0.0092	0.2098	0.3287	-	1.	0.437	
8	SNP_A-2197535	7393	0.4284	0.0799	0.1276	0.2337	0.6047	-	0.565	
9	SNP_A-2299110	7394	0.1711	0.012	0.0185	0.0448	0.1427	0.2584	-	
10										
11										
12										
13										

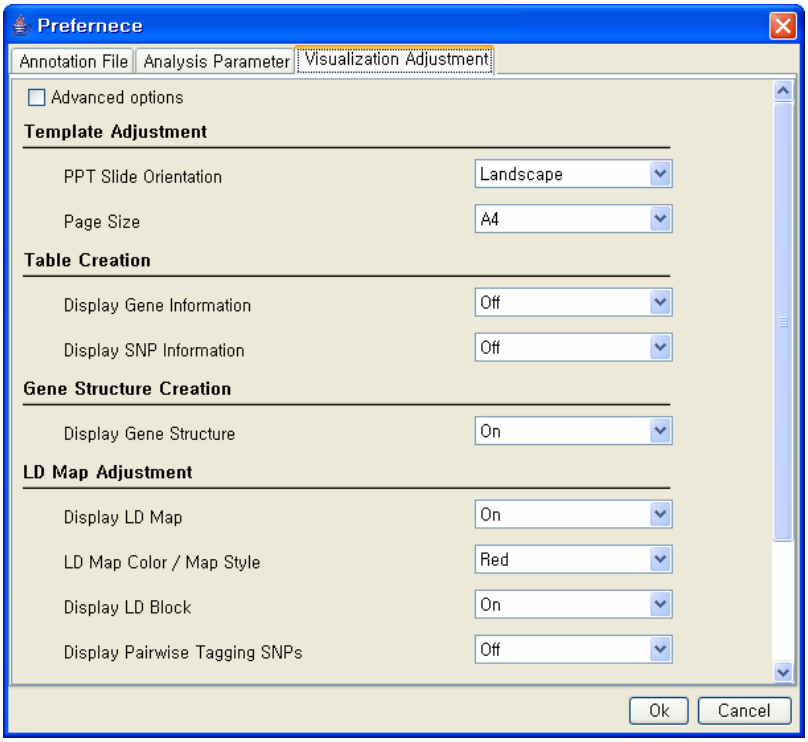
<Figure 2-18> Excel Display

In the <Figure 2-16>, click [Adjust Gene Structure] button, then you will have the interface as seen in the <Figure 2-19>. Detailed control options and usage are identical as describe ahead.



<Figure 2-19> Gene Structure Adjustment

In the <Figure 2-16>, click [Adjust Layout & Parameters] button to change the values of parameters for data analysis and visualization configuration as shown in the <Figure 2-20>. Detailed usage of the interface will be described in **Chapter 3: Parameter Setup**.



<Figure 2-20> Interface of Parameter Configuration

Chapter 3

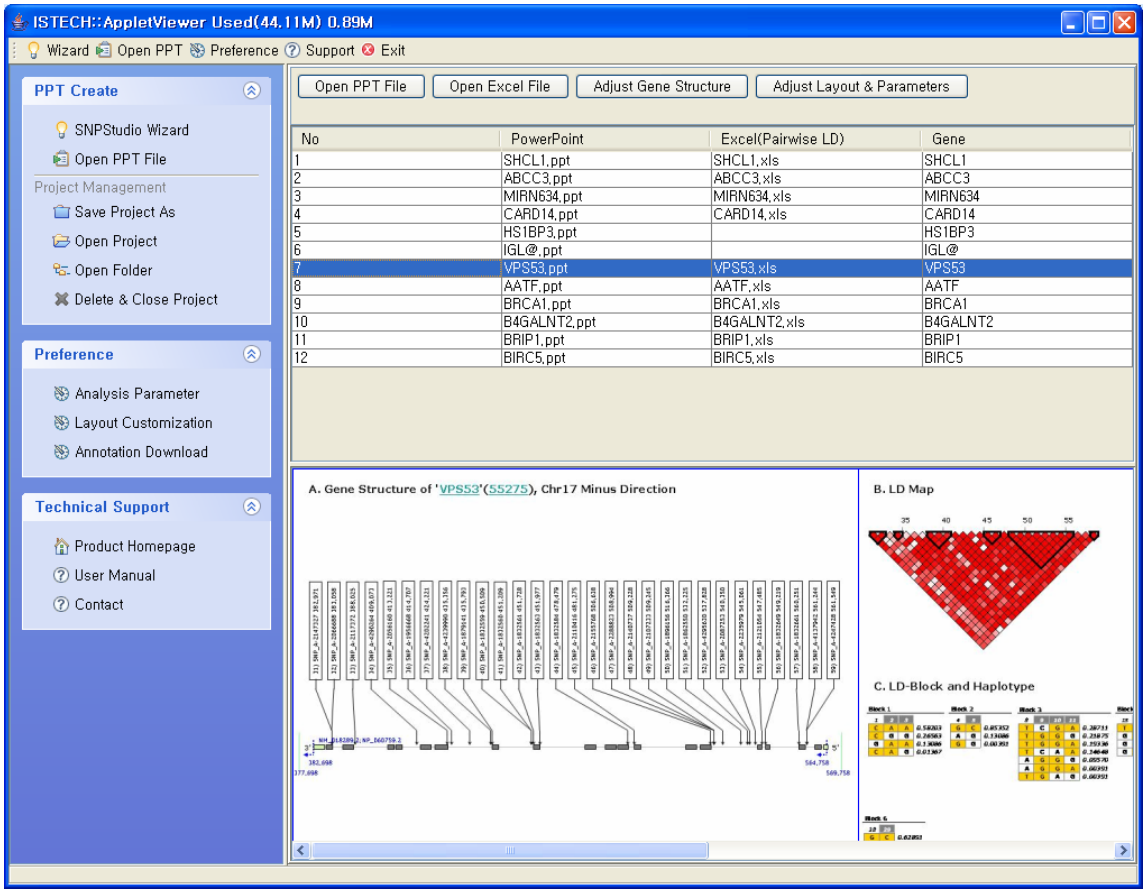
Parameter Setup

3. Parameter Set Up

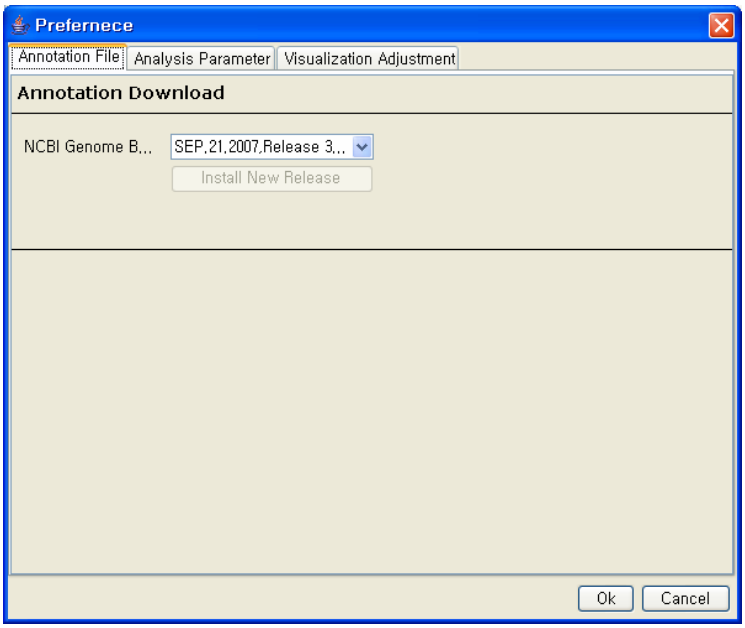
To utilize SNPStudio efficiently, it is possible to modify values of parameters for visualization and data analysis.

3.1. Gene Annotation Information Download and Installation

For gene searching and SNPs mapping, annotation file about genes are required. To download new annotation file, click [Annotation Download] tab in the [Preference] category on the left side of the main interface as you can see in the <Figure 3-1>, and then a window similar to <Figure 3-2> will pops up. Click [Annotation File] tab and select one from the list of annotation files or click [Install New Release] button. Installation of annotation file will proceed after clicking [OK] button.



<Figure 3-1> Main Interface of SNPStudio



<Figure 3-2> Annotation File Download

3.2. Analysis Algorithm Parameter Setup

It will automatically operate the Data Preprocessing and Linkage Disequilibrium Analysis between SNP if there is an Individual Genotype Information in the input data. Followings are the explanations of the Parameter used for the analysis.

- Preprocessing
  - Replace missing genotype with: Replace individual genotype without observed value with one from three possible genotypes (major homozygous genotype, heterozygous genotype, minor homozygous genotype).
  - Flag SNP with missing genotype frequency over: Among observed genotypes, if the ratio of missing genotype is bigger than the fixed value, corresponding SNP will be excluded from the analysis.
  - Flag sample with missing genotype frequency over: Among observed genotypes, if the ratio of missing genotype is bigger than the fixed value, corresponding sample will be excluded from the analysis.
  - Flag SNP with minor allele frequency below: If the observed allele frequency is smaller than the fixed value, corresponding SNP will be excluded from the analysis.
  - Flag SNP by HWE test; p-value below: From the result of Hardy-Weinberg Equilibrium test, if the calculated p-value is smaller than the fixed value, corresponding SNP will be excluded from the analysis.

- HWE p-value multiple correction: Apply Bonferroni multiple test correction.

■ Tagging SNPs Selection

- Minimum allele frequency threshold: If the observed allele frequency is smaller than the fixed value, corresponding SNP is removed from the list of tagging SNPs.
- r^2 threshold: If the square of correlation coefficient (=r) between adjacent SNPs is smaller than the fixed value, corresponding SNPs are removed from the list of tagging SNPs.

■ LD Blocking (Gabriel's method)

- Lower |D'|
- Upper |D'|
- Strong LD fraction
- Minor allele frequency
- Maximum Segment Limit
- Four Gamete Rule: Min. Haplotype Frequency

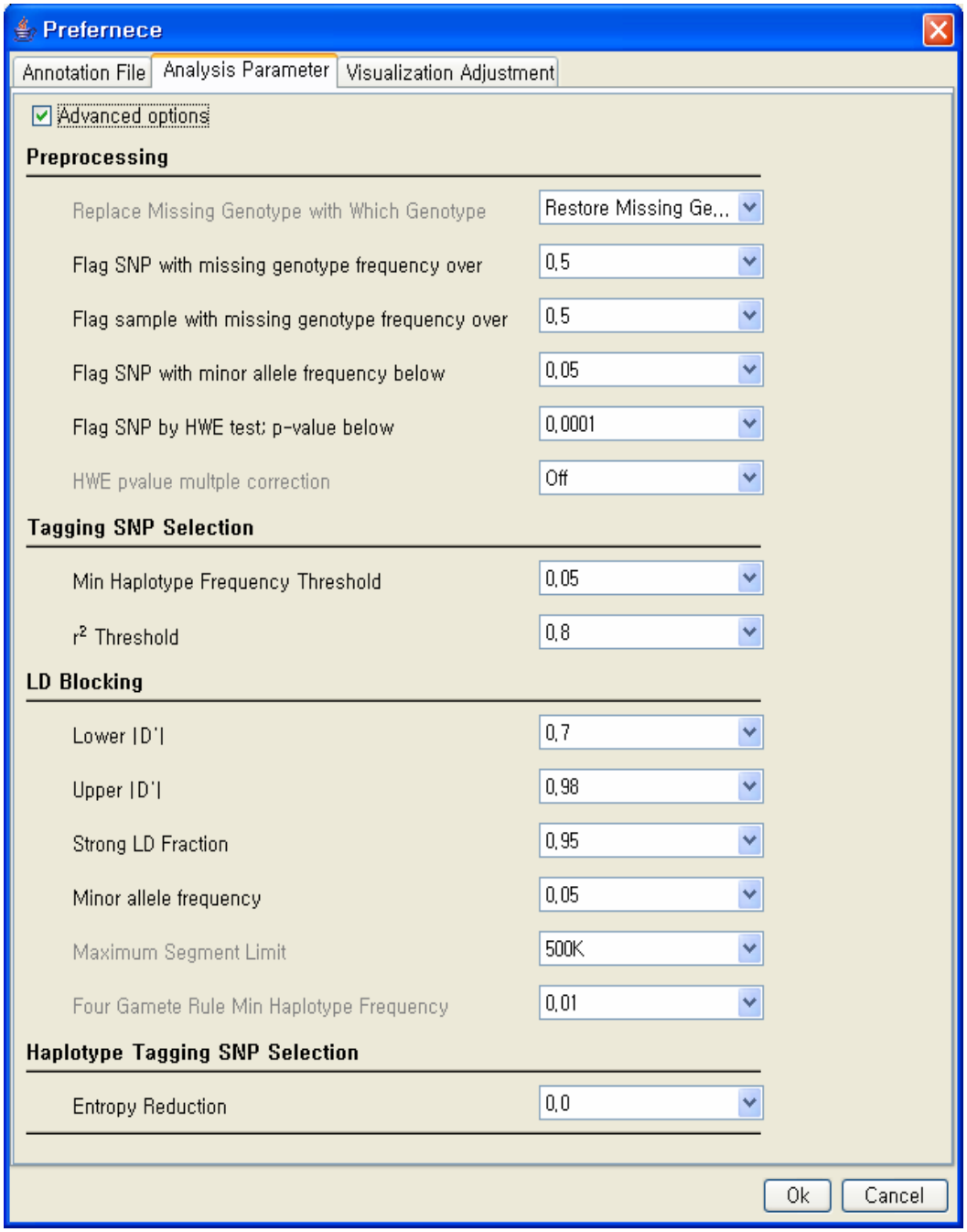
※ Please refer to the related thesis (Gabriel et al, The structure of Haplotype blocks in the human genome. *Science* 2002, 296(5576):2225-2229) for detailed description of the parameters.

■ Haplotype Tagging SNPs Selection

- Entropy Reduction

※ Please refer to the related thesis (Avi-Itzhak et al, Selection of minimum subsets of single nucleotide polymorphisms to capture Haplotype block diversity. *Pac Symp Biocomput* 2003, 466-477) for detailed description of the parameters.

The parameters used for linkage disequilibrium analysis can be set up by clicking [Analysis Parameter] tab from the [Preference] category on the left side of the interface as seen on the <Figure 3-1>. If you want to control hidden parameters, check [Advanced options] as shown in the <Figure 3-3>.



<Figure 3-3> Analysis Parameter Setup

3.3. Visualization Parameter Setup

It is possible to set up parameters used for visualization of SNP and the gene structure with SNP mapping, LD map and etc. as shown in the <Figure 3-4>.

- Template Adjustment

- PPT Slide Orientation: Select whether landscape or portrait
- Page Size: Select whether A3 or A4 size

■ Table Creation

- Display Gene Information: Select “On” to create the gene information related table.
- Display SNP Information: Select “On” to create the SNP information related table.

■ Gene Structure Creation

- Display Gene Structure: Select “On” to visualize gene structure
- Gene Subtitles: It is able to edit/add/remove labels related to the gene using following parameters.
  - {chapter}: identifier of OLE objects (default values are “A”, “B”, “C”, …)
  - {genesym}: gene symbol
  - {geneid}: NCBI gene ID
  - {chrno}: chromosome number
  - {geneorient}: gene orientation
- NCBI Hyperlink on the Gene Symbol: Select “On” to create hyperlink to Entrez Gene DB of NCBI
- NCBI Hyperlink on the Gene ID: Select “On: to create hyperlink to Entrez Gene DB of NCBI
- Height of the Rectangle Representing Exon: Set up the height of exon compared to the default value (ex: 4X means four times the default height)
- Display SNPs over Gene Structure: Select “On” to visualize SNPs locations with the relevant gene
- SNP Subtitles: It is able to edit/add/remove labels related to the SNPs using following parameters.
  - {ridx}: serial number of SNP starting from 1.
  - {snpid}: dbSNP rs ID
  - {chrpos}: position on the chromosome
- Surrounding box for SNP Label: Select “On” to draw box surrounding SNP label
- NCBI Hyperlink on the dbSNP #rs: Select “On” to create hyperlink to dbSNP database of NCBI
- Line Type: Select from off, curved line and folded line
- Rotate SNP Label about 17 degree in right: Select “Off” not to tilt SNP label

■ LD Map Adjustment

- Display LD Map: Select “On” to show linkage disequilibrium pattern in a reverse triangle.
- LD Map Color / Map Style: Six visualization options to change and modify linkage

disequilibrium pattern

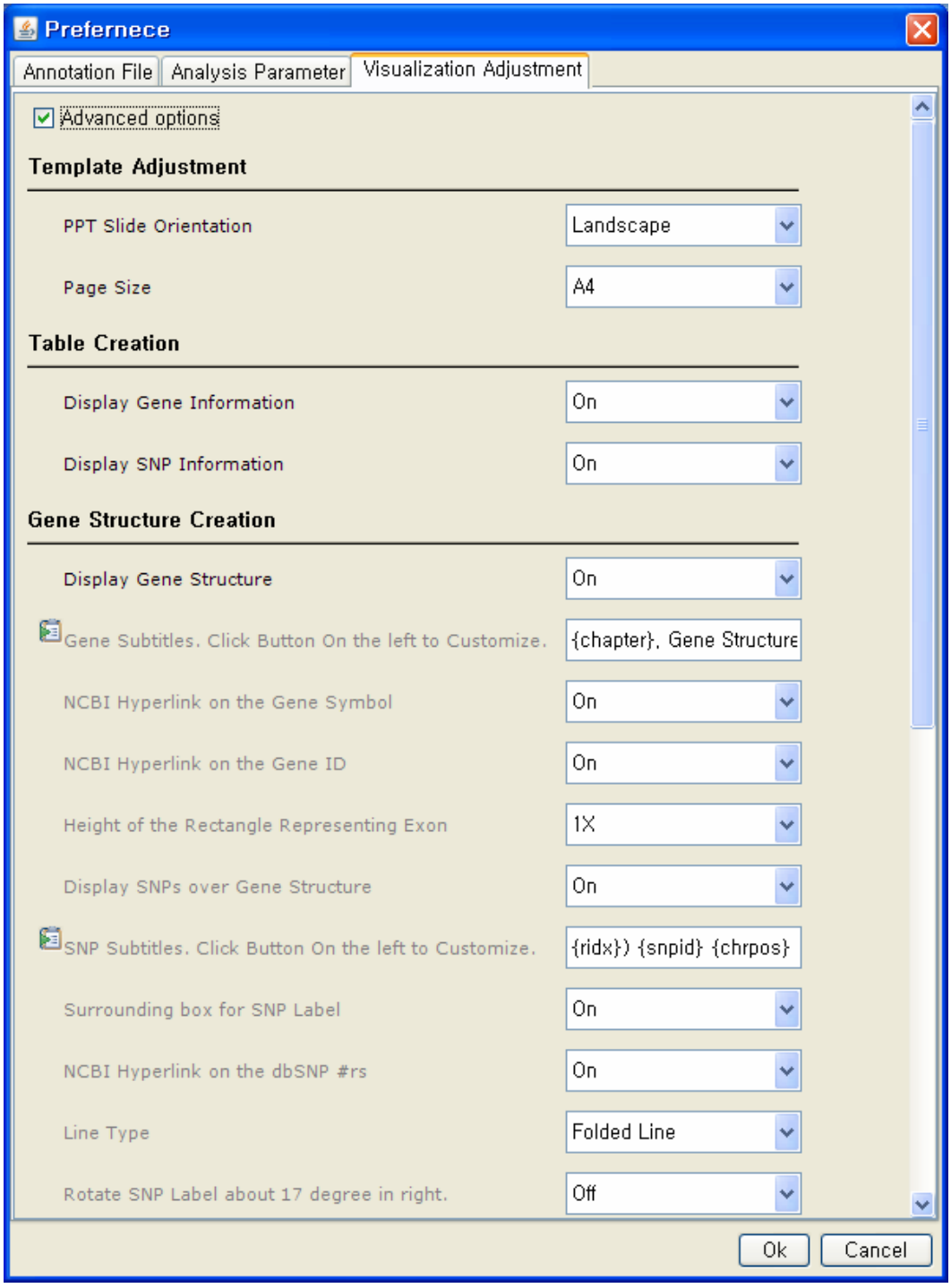
- Display LD Block: Select “On” to visualize LD Blocks
- Display Pairwise Tagging SNPs: Select “On” to visualize Tagging SNPs

■ Haplotype Block Adjustment

- Display Haplotypes in LD Block: Select “On” to show the haplotype information estimated in each LD blocks
- Display Haplotype Tagging SNPs: Select “On” to show haplotype tagging SNPs estimated in each LD blocks

■ LD Matrix Adjustment

- Display LD Matrix: Select “On” to show the calculated |D’| and r^2
- Highlight Strong LD (D-prime): Mark SNP pair showing |D’| is bigger than the set up value with yellow color in LD matrix.
- Highlight Strong LD (r^2): Mark SNP pair showing r^2 is bigger than the set up value with yellow color in LD Matrix
- LD Matrix as Excel File: Select “On” to create calculated LD matrix as Excel file



<Figure 3-4> Visualization Parameter Setup

Chapter 4

Result Interpretation

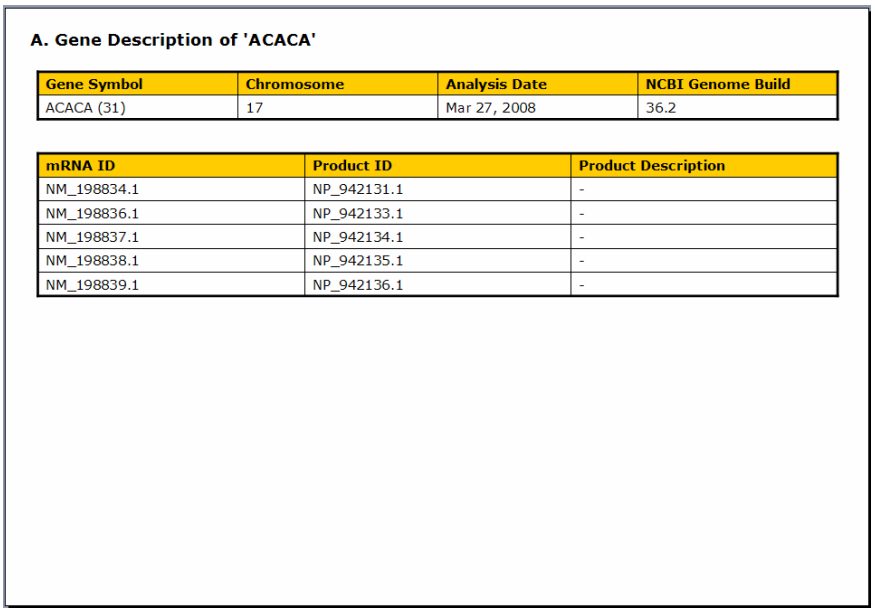
4. Result of Analysis

When the analysis is finished, various results will be created as PowerPoint or Excel file.

4.1. PowerPoint Result

4.1.1. Table Creation

In chapter 3.3, if the [Display Gene Information] is setup as “On” as shown in the <Figure 3-4> and carry out the analysis, basic information of the corresponding gene will be tabulated as shown in the <Figure 4-1>. If the [Display SNP Information] is setup as “On”, the information of the SNPs will be tabulated as shown in the <Figure 4-2>.



A. Gene Description of 'ACACA'

Gene Symbol	Chromosome	Analysis Date	NCBI Genome Build
ACACA (31)	17	Mar 27, 2008	36.2

mRNA ID	Product ID	Product Description
NM_198834.1	NP_942131.1	-
NM_198836.1	NP_942133.1	-
NM_198837.1	NP_942134.1	-
NM_198838.1	NP_942135.1	-
NM_198839.1	NP_942136.1	-

<Figure 4-1> Gene Information Table

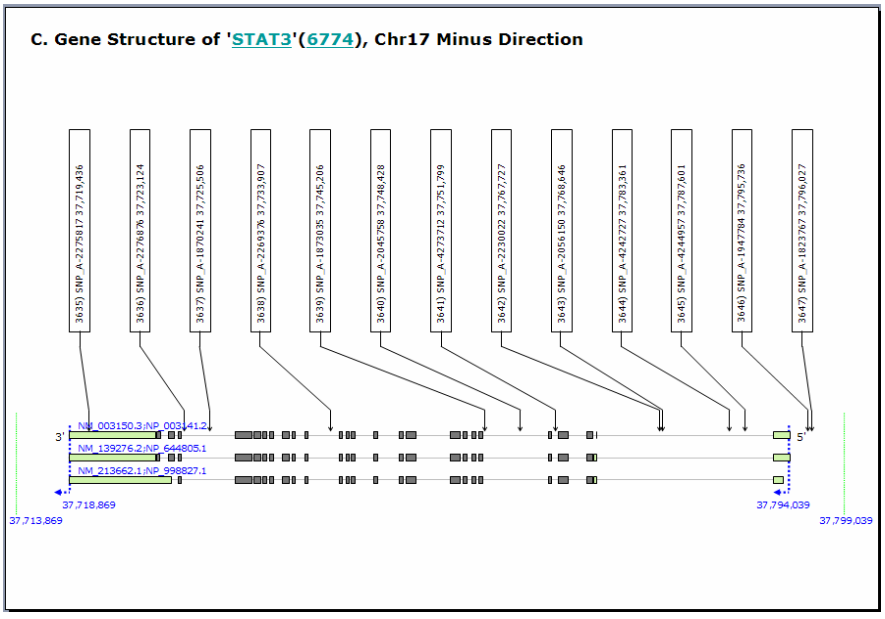
B. SNP List in Gene 'ACACA'

idx	Marker	dbSNP	Gene Pos.
3238	SNP_A-2201671	rs2106772	ACACA+305483
3239	SNP_A-2299557	rs17573357	ACACA+290277
3240	SNP_A-2226795	rs7208415	ACACA+277220
3241	SNP_A-2149796	rs8071315	ACACA+241060
3242	SNP_A-1851108	rs2305098	ACACA+216487
3243	SNP_A-1851114	rs9330250	ACACA+189553
3244	SNP_A-4261196	rs2921368	ACACA+182645
3245	SNP_A-1851117	rs2680398	ACACA+177978
3246	SNP_A-1851118	rs9330251	ACACA+177565
3247	SNP_A-2250320	rs2680399	ACACA+177236
3248	SNP_A-2169728	rs2246459	ACACA+175166
3249	SNP_A-1845112	rs9906044	ACACA+158239
3250	SNP_A-1891105	rs2542655	ACACA+144872
3251	SNP_A-2267887	rs2542660	ACACA+133924
3252	SNP_A-2217843	rs2302803	ACACA+132880
3253	SNP_A-2178139	rs2254914	ACACA+126928
3254	SNP_A-1800500	rs7502022	ACACA+99759
3255	SNP_A-2172396	rs17249662	ACACA+92163
3256	SNP_A-2185150	rs4795190	ACACA+91890
3257	SNP_A-1855707	rs11650464	ACACA+91392
3258	SNP_A-2273163	rs17841479	ACACA+73045
3259	SNP_A-2043365	rs829156	ACACA+49899
3260	SNP_A-4298658	rs1473626	ACACA+45110
3261	SNP_A-2239233	rs11867736	ACACA+16526

<Figure 4-2> SNP Information Table

4.1.2. SNP Location and Gene Structure Creation

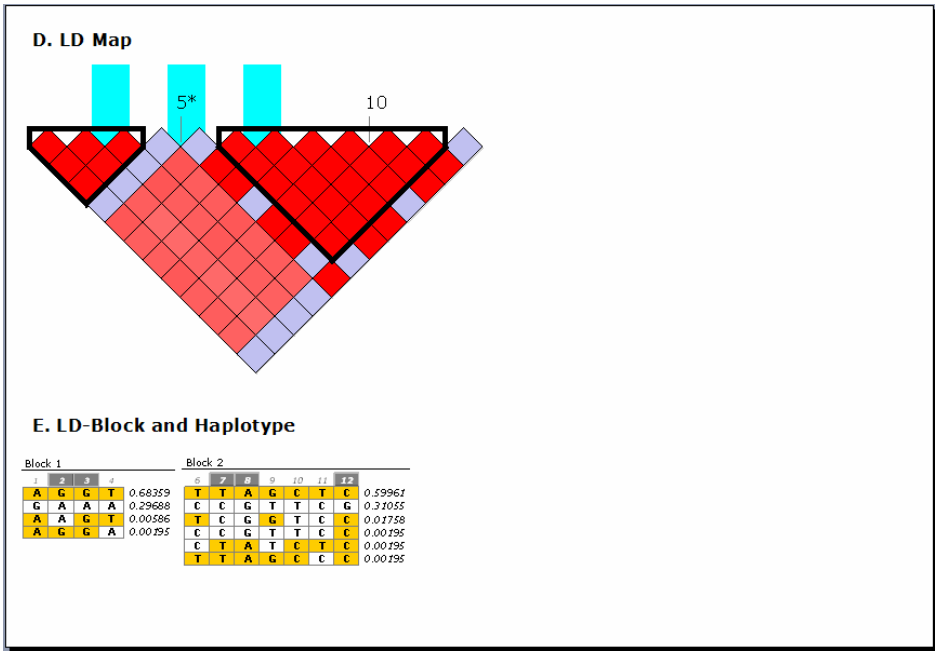
<Figure 4.3> is the PowerPoint slide showing SNPs and gene structure. Gene symbol and gene ID are hyperlinked to the Entrez Gene DB of NCBI. The rectangles in the middle of the slide represent SNPs and linked to the gene according to their chromosome positions with arrow line. The RNA structures of gene are depicted with exon, intron and UTR. Green color means UTR and gray color means coding region. Please refer to the **Chapter 3.3** for detailed control of the visualization.



<Figure 4-3> SNP Location and Gene Structure

4.1.3. LD Map Adjustment

If genotype information is available, visualization of the results of linkage disequilibrium (LD) analysis is implemented automatically as shown in the <Figure 4-4>. Strong red color means that there is strong linkage disequilibrium between adjacent SNPs. The light blue rectangle on the top of the linkage disequilibrium map is showing tagging SNP. LD block is surrounded by thick black line. Haplotypes and their frequencies estimated in each LD blocks are tabulated in the bottom of the slide. Yellow color is showing major allele of the relevant SNP. The haplotype tagging SNP is marked with gray box in the haplotype table. For more information, please refer to **Chapter 3.3**.



<Figure 4-4> LD Map and Haplotypes

4.1.4. LD Matrix Adjustment

<Figure 4-5> shows linkage disequilibrium matrix calculated using individuals' genotypes. As linkage disequilibrium indices, |D'| and r^2 values are calculated. |D'| values and r^2 values are displayed in the upper triangle and lower triangle separately. Complete linkage disequilibrium (i.e. |D'|=1 and r^2=1) relationships are colored in yellow as shown in the <Figure 4-5>. For more detailed information, please refer to **Chapter 3.3**.

F. LD Matrix

r2 \ |D'|	1	2	3	4	5	6	7	8	9	10	11	12	13
1		1	1	1	1	0.807	0.789	0.789	0.8	0.789	0.79	0.781	1
2	0.972		1	0.991	1	0.791	0.763	0.763	0.776	0.763	0.763	0.758	1
3	1	0.973		1	1	0.809	0.79	0.79	0.802	0.79	0.791	0.783	1
4	0.991	0.963	0.991		1	0.811	0.793	0.793	0.804	0.793	0.794	0.785	1
5	0.01	0.011	0.01	0.01		1	1	1	1	1	1	1	1
6	0.566	0.555	0.569	0.578	0.01		0.991	0.991	1	0.991	0.99	1	1
7	0.492	0.475	0.496	0.503	0.044	0.906		1	0.991	1	1	1	1
8	0.492	0.475	0.496	0.503	0.044	0.906	1		0.991	1	1	1	1
9	0.559	0.541	0.562	0.57	0.012	1	0.892	0.892		0.991	0.99	1	1
10	0.492	0.475	0.496	0.503	0.044	0.906	1	1	0.892		1	1	1
11	0.485	0.467	0.488	0.496	0.041	0.903	0.991	0.991	0.888	0.991		1	1
12	0.547	0.529	0.55	0.558	0.012	0.982	0.892	0.892	0.982	0.892	0.889		1
13	0.01	0.011	0.01	0.01	1	0.01	0.044	0.044	0.012	0.044	0.041	0.012	

<Figure 4-5> LD Matrix

Chapter 5

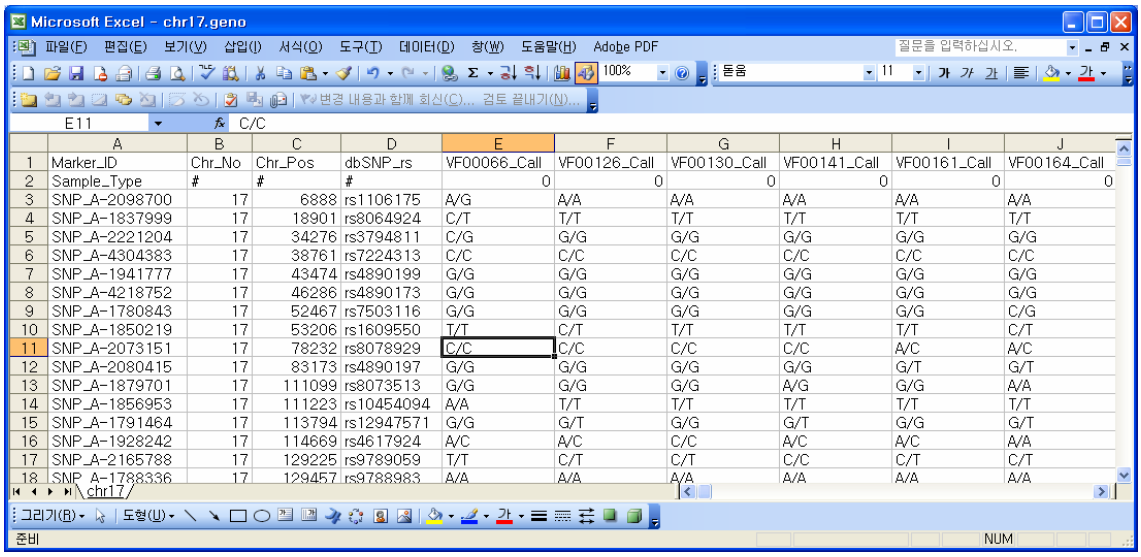
Input Data

5. Input Data Format

5.1. SNP Analyzer-Pro Format

This is the tab-delimited text file format. The first line and the second line are used as headers which are followed by the SNP information and individual genotypes. <Figure 5-1> shows an example and details are as follows.

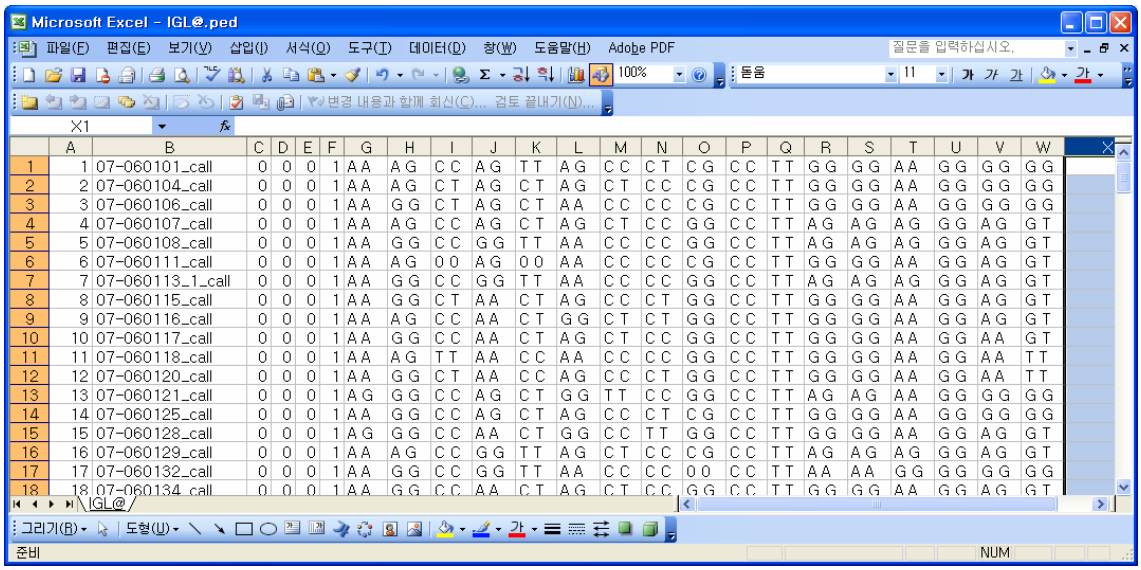
- The First Row
  - The First Column (Marker_ID): This is to classify SNP. It is a mandatory reserved word.
  - The Second Column (Chr_No): This is the chromosome number of SNP. It is a mandatory reserved word.
  - The Third Column (Chr_Pos): This is the location of SNP on the chromosome. It is a mandatory reserved word.
  - The Fourth Column (dbSNP_rs): This is the fixed SNP ID in the NCBI dbSNP database. It is a mandatory reserved word.
  - The Remaining Columns: These are individual IDs. If there is no genotype information, there is no need to describe.
- The Second Row
  - The First Column (Sample_Type): This is for the next version of SNPStudio. It is a mandatory reserved word.
  - The Second ~ The Fourth Column: It is marked with "#". It is a mandatory reserved word.
  - The Remaining Columns: These are for the next version of SNPStudio, and marked in "0" or "1". If there is no genotype information, these are not necessary.
- The Remaining Rows: These are the actual values corresponding to each columns of the first row.



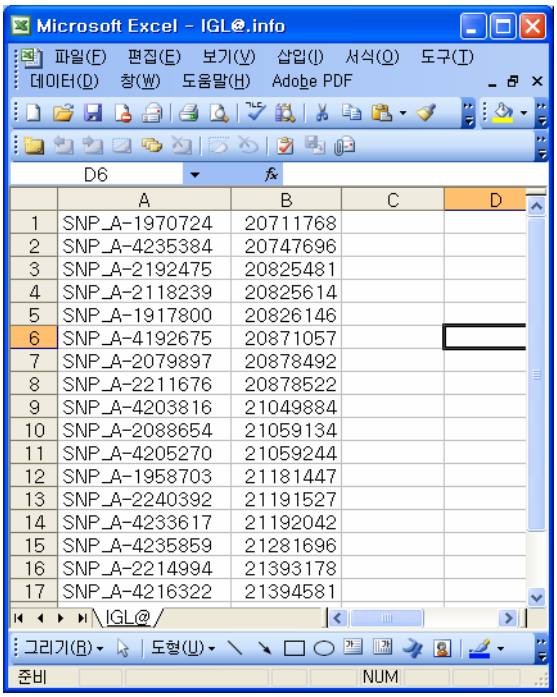
<Figure 5-1> SNPAnalyzer-Pro Format

5.2. Haploview / PLINK Format

<Figure 5-2> is an example of input data for Haploview and PLINK Program. In case of Haploview, it is the Standard Linkage Format. To locate SNPs on genes, annotation file with SNP location information is needed separately along with genotype file.



<Figure 5-2> Standard Linkage Format (Haploview Genotype)



<Figure 5-3> SNP Annotation (Haploview)

5.3. International HapMap Genotype Format

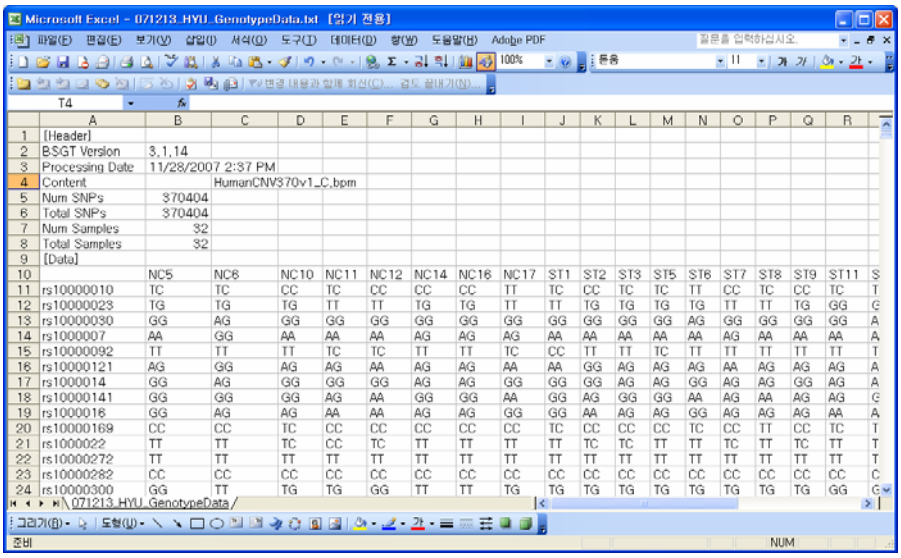
<Figure 5-4> is an example of International HapMap genotype data.

```
#Fri Mar 14 01:43:29 2008: HapMap genotype data dump, SNPs genotyped in population CEU on chr17:32
#For details on file format, see http://www.hapmap.org/genotypes/
rs# alleles chrom pos strand assembly# center protLSID assayLSID panelLSID QCcode NA06985 NA06991
rs9892334 A/G chr17 32381183 + ncbi_B36 affymetrix urn:LSID:affymetrix.hapmap.org:Protocol:genotyp
rs8077789 C/G chr17 32383839 + ncbi_B36 imsut-riken urn:lsid:imsut-riken.hapmap.org:Protocol:genot
rs10908286 A/C chr17 32383961 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping
rs11650575 A/C chr17 32384701 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping
rs8067751 A/C chr17 32384744 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping_
rs12051731 C/G chr17 32384818 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping
rs4408589 C/G chr17 32385876 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping_
rs7218038 A/C chr17 32389341 + ncbi_B36 affymetrix urn:LSID:affymetrix.hapmap.org:Protocol:genotyp
rs6607358 A/G chr17 32389367 + ncbi_B36 illumina urn:LSID:illumina.hapmap.org:Protocol:Infinium_ge
rs11871752 C/T chr17 32389923 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping
rs9895466 A/G chr17 32390954 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping_
rs9972935 A/G chr17 32391264 + ncbi_B36 imsut-riken urn:lsid:imsut-riken.hapmap.org:Protocol:genot
rs9893214 C/T chr17 32392031 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping_
rs11871466 A/G chr17 32392482 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping
rs4427852 A/T chr17 32392660 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping_
rs12942174 A/G chr17 32393071 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping
rs7222936 C/T chr17 32393438 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping_
rs8069751 A/G chr17 32394215 + ncbi_B36 perlegen urn:lsid:perlegen.hapmap.org:Protocol:Genotyping
```

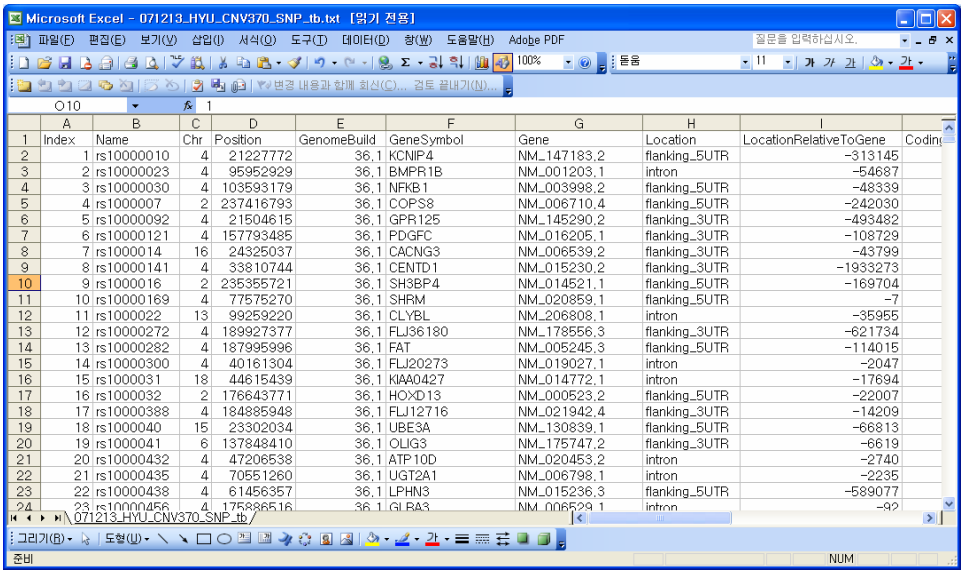
<Figure 5-4> Internal HapMap Genotype Format

5.4. Illumina Golden Gate Format

It is able to use GoldenGate format data of Illumina as input file. The <Figure 5-5> is an example of the genotype result data extracted using “matrix format” in the BeadStudio program provided from Illumina. The <Figure 5-6> is an example of the result data by saving “SNP Report Table” displayed in the interface of BeadStudio program.



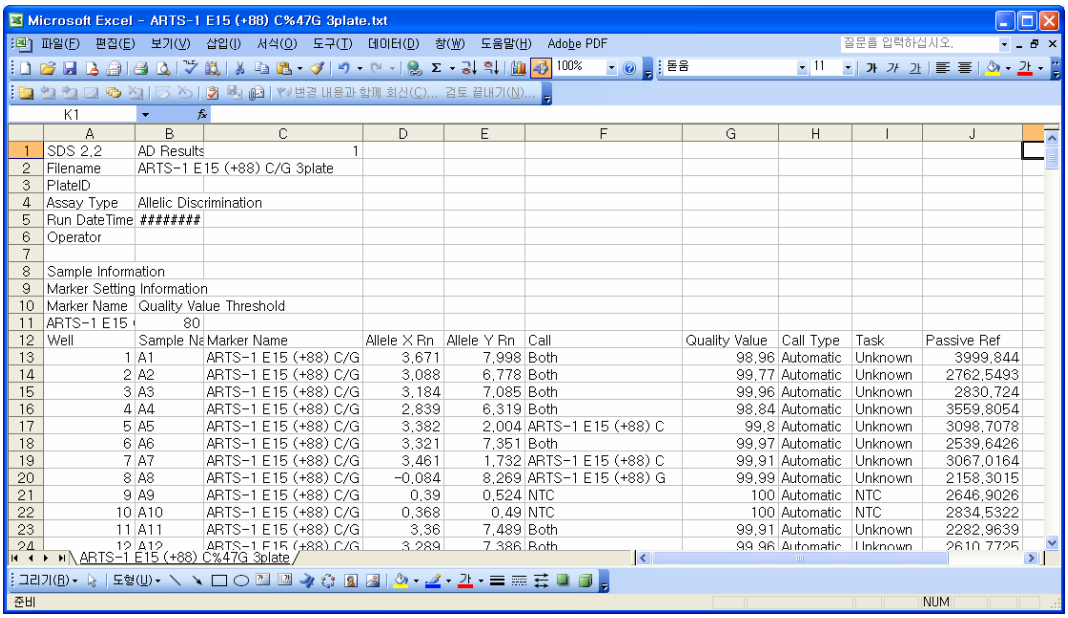
<Figure 5-5> GoldenGate Genotype Format



<Figure 5-6> BeadStudio SNP Table (Golden Gate)

5.5. ABI TaqMan Format

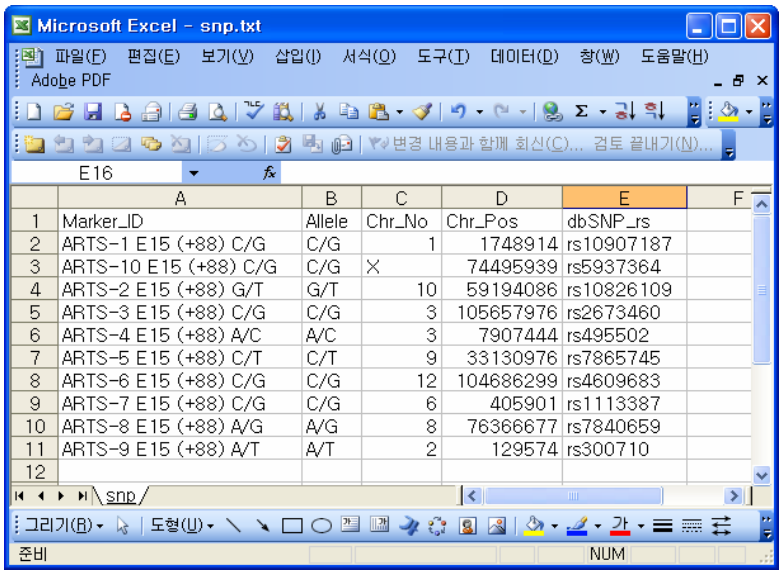
<Figure 5-7> is an example of genotype data and <Figure 5-8> is an example of SNP information data created from TaqMan of ABI.



Microsoft Excel - ARTS-1 E15 (+88) C%47G 3plate.txt

	A	B	C	D	E	F	G	H	I	J
1	SDS 2.2	AD Results	1							
2	Filename	ARTS-1 E15 (+88) C/G 3plate								
3	PlateID									
4	Assay Type	Allelic Discrimination								
5	Run DateTime	########								
6	Operator									
7										
8	Sample Information									
9	Marker Setting Information									
10	Marker Name	Quality Value Threshold								
11	ARTS-1 E15	80								
12	Well	Sample Na	Marker Name	Allele X Rn	Allele Y Rn	Call	Quality Value	Call Type	Task	Passive Ref
13	1	A1	ARTS-1 E15 (+88) C/G	3,671	7,998	Both	98,96	Automatic	Unknown	3999,844
14	2	A2	ARTS-1 E15 (+88) C/G	3,088	6,778	Both	99,77	Automatic	Unknown	2762,5493
15	3	A3	ARTS-1 E15 (+88) C/G	3,184	7,085	Both	99,96	Automatic	Unknown	2830,724
16	4	A4	ARTS-1 E15 (+88) C/G	2,839	6,319	Both	98,84	Automatic	Unknown	3559,8054
17	5	A5	ARTS-1 E15 (+88) C/G	3,382	2,004	ARTS-1 E15 (+88) C	99,8	Automatic	Unknown	3098,7078
18	6	A6	ARTS-1 E15 (+88) C/G	3,321	7,351	Both	99,97	Automatic	Unknown	2539,6426
19	7	A7	ARTS-1 E15 (+88) C/G	3,461	1,732	ARTS-1 E15 (+88) C	99,91	Automatic	Unknown	3067,0164
20	8	A8	ARTS-1 E15 (+88) C/G	-0,084	8,269	ARTS-1 E15 (+88) G	99,99	Automatic	Unknown	2158,3015
21	9	A9	ARTS-1 E15 (+88) C/G	0,39	0,524	NTC	100	Automatic	NTC	2646,9026
22	10	A10	ARTS-1 E15 (+88) C/G	0,368	0,49	NTC	100	Automatic	NTC	2834,5322
23	11	A11	ARTS-1 E15 (+88) C/G	3,36	7,489	Both	99,91	Automatic	Unknown	2282,9639
24	12	A12	ARTS-1 E15 (+88) C/G	3,289	7,386	Both	99,96	Automatic	Unknown	2610,7725

<Figure 5-7> Genotype of ABI TaqMan



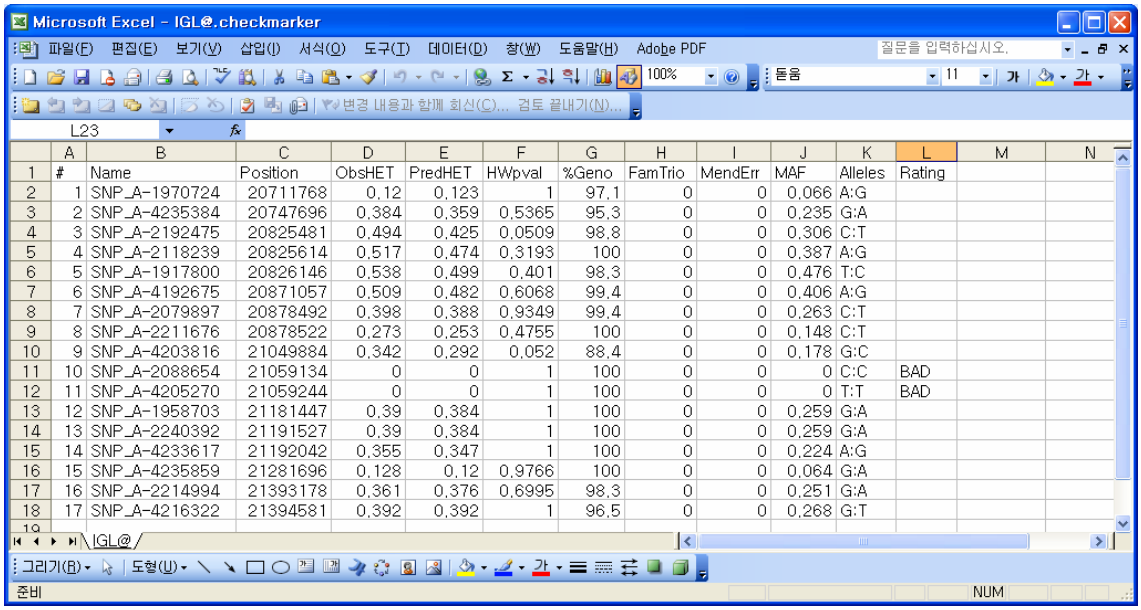
Microsoft Excel - snp.txt

	A	B	C	D	E	F
1	Marker_ID	Allele	Chr_No	Chr_Pos	dbSNP_rs	
2	ARTS-1 E15 (+88) C/G	C/G	1	1748914	rs10907187	
3	ARTS-10 E15 (+88) C/G	C/G	X	74495939	rs5937364	
4	ARTS-2 E15 (+88) G/T	G/T	10	59194086	rs10826109	
5	ARTS-3 E15 (+88) C/G	C/G	3	105657976	rs2673460	
6	ARTS-4 E15 (+88) A/C	A/C	3	7907444	rs495502	
7	ARTS-5 E15 (+88) C/T	C/T	9	33130976	rs7865745	
8	ARTS-6 E15 (+88) C/G	C/G	12	104686299	rs4609683	
9	ARTS-7 E15 (+88) C/G	C/G	6	405901	rs1113387	
10	ARTS-8 E15 (+88) A/G	A/G	8	76366677	rs7840659	
11	ARTS-9 E15 (+88) A/T	A/T	2	129574	rs300710	
12						

<Figure 5-8> Marker Information of ABI TaqMan

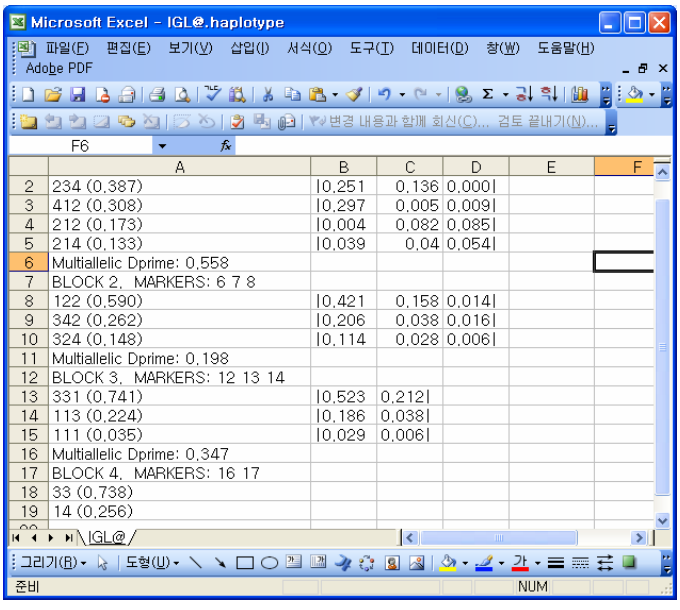
5.6. Result Files from Haploview

The output of Haploview program can be used in the SNPStudio. <Figure 5-9> is an example of “Marker Information Result” of Haploview program. <Figure 5-10>, <Figure 5-11>, <Figure 5-12> is each an example of “Haplotype Result”, “LD Result” and “Tagging SNPs” of Haploview program.


Microsoft Excel - IGL@.checkmarker

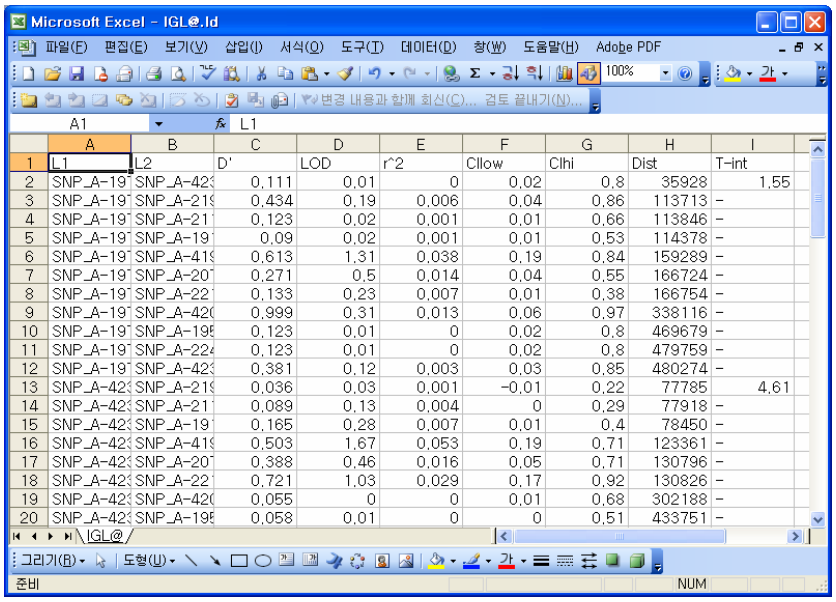
#	Name	Position	ObsHET	PredHET	HWpval	%Geno	FamTrio	MendErr	MAF	Alleles	Rating
1	SNP_A-1970724	20711768	0,12	0,123	1	97,1	0	0	0,066	A:G	
2	SNP_A-4235384	20747696	0,384	0,359	0,5365	95,3	0	0	0,235	G:A	
3	SNP_A-2192475	20825481	0,494	0,425	0,0509	98,8	0	0	0,306	C:T	
4	SNP_A-2118239	20825614	0,517	0,474	0,3193	100	0	0	0,387	A:G	
5	SNP_A-1917800	20826146	0,538	0,499	0,401	98,3	0	0	0,476	T:C	
6	SNP_A-4192675	20871057	0,509	0,482	0,6068	99,4	0	0	0,406	A:G	
7	SNP_A-2079897	20878492	0,398	0,388	0,9349	99,4	0	0	0,263	C:T	
8	SNP_A-2211676	20878522	0,273	0,253	0,4755	100	0	0	0,148	C:T	
9	SNP_A-4203816	21049884	0,342	0,292	0,052	88,4	0	0	0,178	G:C	
10	SNP_A-2088654	21059134	0	0	1	100	0	0	0	C:C	BAD
11	SNP_A-4205270	21059244	0	0	1	100	0	0	0	T:T	BAD
12	SNP_A-1958703	21181447	0,39	0,384	1	100	0	0	0,259	G:A	
13	SNP_A-2240392	21191527	0,39	0,384	1	100	0	0	0,259	G:A	
14	SNP_A-4233617	21192042	0,355	0,347	1	100	0	0	0,224	A:G	
15	SNP_A-4235859	21281696	0,128	0,12	0,9766	100	0	0	0,064	G:A	
16	SNP_A-2214994	21393178	0,361	0,376	0,6995	98,3	0	0	0,251	G:A	
17	SNP_A-4216322	21394581	0,392	0,392	1	96,5	0	0	0,268	G:T	

<Figure 5-9> Marker Information Result of Haploview

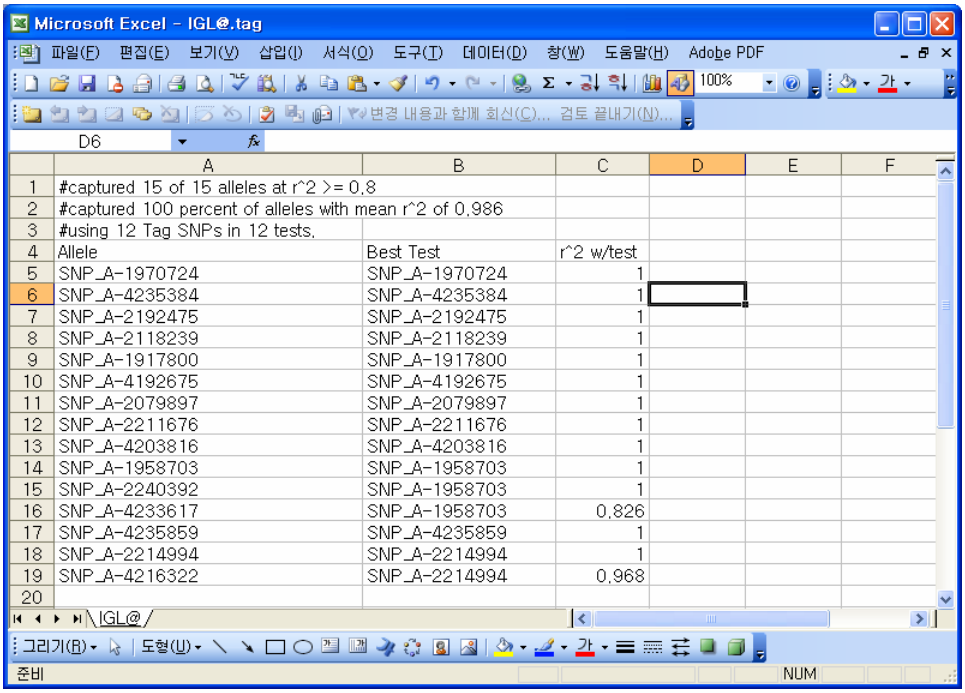

Microsoft Excel - IGL@.haplotype

234 (0,387)	|0,251 0,136 0,000|
412 (0,308)	|0,297 0,005 0,009|
212 (0,173)	|0,004 0,082 0,085|
214 (0,133)	|0,039 0,04 0,054|
Multiallelic Dprime: 0,558	
BLOCK 2, MARKERS: 6 7 8	
122 (0,590)	|0,421 0,158 0,014|
342 (0,262)	|0,206 0,038 0,016|
324 (0,148)	|0,114 0,028 0,006|
Multiallelic Dprime: 0,198	
BLOCK 3, MARKERS: 12 13 14	
331 (0,741)	|0,523 0,212|
113 (0,224)	|0,186 0,038|
111 (0,035)	|0,029 0,006|
Multiallelic Dprime: 0,347	
BLOCK 4, MARKERS: 16 17	
33 (0,738)	
14 (0,256)	

<Figure 5-10> Haplotype Result of Haploview



<Figure 5-11> LD Result of Haploview



	A	B	C	D	E	F
1	#captured 15 of 15 alleles at r^2 >= 0.8					
2	#captured 100 percent of alleles with mean r^2 of 0.986					
3	#using 12 Tag SNPs in 12 tests.					
4	Allele	Best Test	r^2 w/test			
5	SNP_A-1970724	SNP_A-1970724	1			
6	SNP_A-4235384	SNP_A-4235384	1			
7	SNP_A-2192475	SNP_A-2192475	1			
8	SNP_A-2118239	SNP_A-2118239	1			
9	SNP_A-1917800	SNP_A-1917800	1			
10	SNP_A-4192675	SNP_A-4192675	1			
11	SNP_A-2079897	SNP_A-2079897	1			
12	SNP_A-2211676	SNP_A-2211676	1			
13	SNP_A-4203816	SNP_A-4203816	1			
14	SNP_A-1958703	SNP_A-1958703	1			
15	SNP_A-2240392	SNP_A-1958703	1			
16	SNP_A-4233617	SNP_A-1958703	0.826			
17	SNP_A-4235859	SNP_A-4235859	1			
18	SNP_A-2214994	SNP_A-2214994	1			
19	SNP_A-4216322	SNP_A-2214994	0.968			
20						

<Figure 5-12> Tagging SNPs of Haploview